

2021 Seminars

12/01/2021

"Predicting future risk of colorectal cancer in patients with ulcerative colitis"

[Kathleen \(Kit\) Curtius, PhD](#)

1:00 PM – 2:00 PM PST

ABSTRACT: Patients with ulcerative colitis (UC) have a 4-fold increased risk of developing colorectal cancer than the general population and thus are advised to participate in regular surveillance to remove pre-cancers that may be detected during colonoscopy. We developed and validated a statistical model for predicting patient-specific risk of progression over time. With this model, we created a translational web-tool called UC-CaRE (Ulcerative Colitis-Cancer Risk Estimator) that can be used to calculate and communicate individualized cancer risk estimates to UC patients with low-grade dysplasia based on their clinico-pathological features. Using low-cost shallow whole genome sequencing in retrospective case-control studies, we have also found that copy number alterations in UC predict future neoplastic risk in patients with high accuracy. As molecular-based decision-making becomes more prominent in the clinical setting of early cancer detection, we propose that models of genotype evolution can be integrated into tools like UC-CaRE to truly personalize care.

11/03/2021

"Design and analysis considerations for an umbrella trial for acute myeloid leukemia"

[Megan Othus, PhD](#)

1:00 PM – 2:00 PM PST

ABSTRACT: Acute myeloid leukemia (AML) is a rare cancer of the bone marrow with very poor outcomes for many patients. Mutations that drive the biology and prognosis of AML are well characterized and therapies targeting these mutations are being rapidly developed. Measurable residual disease (MRD) is the strongest prognostic factor for AML patients and clinicians want to be able to use it to make treatment decisions, but there is not consensus on how to define MRD and how and when to use MRD to guide treatment decisions. myeloMATCH is an NCI-funded umbrella trial being developed for patients with AML with the goals to (1) evaluate both targeted and non-targeted agents in the treatment of AML and (2) to evaluate the role of MRD in AML. This talk will discuss the structure and design considerations for the screening portion of the myeloMATCH trial and also discuss some of the challenges with evaluating MRD within and across clinical trials part of myeloMATCH.

10/06/2021

"Data driven analytical approaches informing the UCSD Return to Learn program"

[Natasha Martin, DPhil](#)

1:00 PM – 2:00 PM PST

ABSTRACT: The UC San Diego Return to Learn program was founded on a data-driven, adaptive strategy designed to improve campus safety during the COVID-19 pandemic. Dr. Martin will discuss some of the data-driven analytical approaches used to inform the RtL program in relation to wastewater monitoring and asymptomatic testing frequency.

05/05/2021

"Surrogate Assisted Semi-Supervised Inference for High Dimensional Risk Prediction"

[Jue "Marquis" Hou, PhD](#)

1:00 PM – 2:00 PM PST

BACKGROUND: Precise risk prediction is vitally important for clinical care. High risk patients can be monitored more intensively. More aggressive treatment options should be considered for those unlikely to respond to the standard therapy. Developing generalizable precision medicine strategies requires real world data (RWD). A valuable source of RWD is the electronic health records (EHR) which accrue detailed clinical data in broad patient populations. Risk modeling with EHR data, however, is challenging due to a lack of direct observations on the disease outcome, Y , and the high dimensionality of the candidate predictors, X . In this paper, we develop a surrogate assisted semi-supervised-learning (SAS) approach to risk modeling with high dimensional predictors, leveraging a large size N of unlabeled observations on X and surrogates of Y , S , as well as a small size n labeled observations on X , S , and Y . The SAS procedure borrows information from S along with X to impute the unobserved Y via a sparse working imputation model with moment conditions to achieve robustness against misspecification in the imputation model and a one-step bias correction to enable interval estimation for the predicted risk. We demonstrate that the SAS procedure provides valid inference for the predicted risk derived from a high dimensional working model, $P(Y = 1|X)$, even when the underlying population parameter for is dense and the risk model is mis-specified. We present an extensive simulation study to demonstrate the superiority of our SSL approach compared to existing supervised methods. We apply the method to derive genetic risk prediction of type-2 diabetes mellitus using a EHR biobank cohort.

04/07/2021

"Proximal Causal Learning for Complex Longitudinal Studies"

[Andrew Ying, PhD](#)

1:00 PM – 2:00 PM PST

BACKGROUND: A standard assumption for causal inference with longitudinal data is that at each follow-up time, one has measured a sufficiently rich set of covariates to ensure that within covariate strata, subjects are exchangeable across observed treatment values, also known as "sequential randomization assumption (SRA)". Skepticism about SRA in observational studies is often warranted because it hinges on investigators' ability to accurately measure covariates over time capturing all potential sources of time-varying confounding. Realistically, confounding mechanisms can rarely if ever, be learned with certainty from measured covariates. One can therefore only ever hope that covariate measurements are at best proxies of true underlying confounding mechanisms operating in an observational study, thus invalidating causal claims made on basis of SRA. In this paper, we extend the proximal causal inference framework of Tchetgen Tch-etgen, Ying, et al. (2020) and Cui et al. (2020) to the longitudinal setting under a semiparametric marginal structural mean model (MSMM). The longitudinal proximal inference approach we propose offers an opportunity to learn about joint causal effects in settings where SRA on the basis of measured time-varying covariates fails, by formally accounting for the covariate measurements as imperfect proxies of underlying confounding mechanisms. We establish sufficient conditions for nonparametric identification with the aid of a pair of time varying proxies, when sequential randomization fails to hold due to unmeasured confounding. We provide a characterization of all regular and asymptotically linear estimators of the parameter indexing the MSMM, including a rich class of doubly robust estimators, and establish the corresponding semiparametric efficiency bound for the MSMM. Our approach is illustrated via extensive simulation studies and a data application on potential protective effects of the anti-rheumatic therapy Methotrexate (MTX) among patients with rheumatoid arthritis.

02/03/2021

"Statistical Consulting: The Unplanned Road"

Joni Ricks-Oddie, PhD, MPH

1:00 PM – 2:00 PM PST

BACKGROUND: Dr. Ricks-Oddie is currently the Director of the University of California, Irvine Center for Statistical Consulting in the Department of Statistics and the Biostatistics, Epidemiology and Research Design Unit in the Institute for Clinical and Translational Science. Dr. Ricks-Oddie oversees general operations and strategic planning for the both units. She will be speaking regarding her experience as a statistical consultant in an academic environment. Statistical consulting is a dynamic and challenging environment that allows the consultant to work on projects from a wide range of backgrounds. As a consultant you have the opportunity to work on real world projects and imperfect data with unfamiliar subject matter. The nature of the position requires constant professional development and is a great way for new statisticians to build their network and their skills.

01/06/2021

"Learning survival from EMR/EHR data to estimate treatment effects using high dimensional claims codes"

Lily (Ronghui) Xu, PhD

1:00 PM – 2:00 PM PST

BACKGROUND: Our work was motivated by the analysis projects using the linked US SEER-Medicare database to study treatment effects in men of age 65 years or older who were diagnosed with prostate cancer. Such data sets contain up to 100,000 human subjects and over 20,000 claim codes. The data were obviously not randomized with regard to the treatment of interest, for example, radical prostatectomy versus conservative treatment. Informed by previous instrumental variable (IV) analysis, we know that confounding most likely exists beyond the commonly captured clinical variables in the database, and meanwhile the high dimensional claims codes have been shown to contain rich information about the patients' survival. Hence we aim to incorporate the high dimensional claims codes into the estimation of the treatment effect. The orthogonal score method is one that can be used for treatment effect estimation and inference despite the bias induced by regularization under the high dimensional hazards outcome model and the high dimensional treatment model. In addition, we show that with cross-fitting the approach has rate doubly-robust property in high dimensions.